

KEKER
VAN NEST
& PETERS

LATHAM & WATKINS LLP

MORRISON FOERSTER

June 10, 2025

Hon. Ona T. Wang
Daniel Patrick Moynihan
United States Courthouse
500 Pearl St.
New York, NY 10007-1312

cc: All Counsel of Record (via ECF)

Re: Discovery dispute concerning preservation The Times's records of its use of OpenAI's LLMs

Dear Judge Wang:

OpenAI requests a pre-motion conference regarding The Times's refusal to address whether it has preserved data from [REDACTED]—a data repository that logs prompts and outputs from Times employees' use of OpenAI's API. The Times has agreed to produce evidence of its use of OpenAI's generative AI ("GenAI") technology generally and to produce certain [REDACTED] data specifically. The Times assured OpenAI that such data is preserved [REDACTED].

But, in April, The Times revealed it has virtually zero [REDACTED] logs for the employees focused on its GenAI initiatives. OpenAI has repeatedly asked The Times to explain this apparent discrepancy, but The Times has refused to (a) promptly explain why it has not retained logs for these individuals or (b) reveal the full extent of this potential preservation issue.

The possible failure to preserve this data is troubling, given The Times's concession of its relevance, and The Times's refusal to provide clarity is impeding the ultimate production of this data. OpenAI therefore requests the Court order The Times, by July 2, to: (1) confirm it is following its preservation practices for [REDACTED] data related to relevant Times employees; (2) provide counts for prompts made by those employees; and (3) provide complete responses to a discrete set of preservation questions.

1. Background.

[REDACTED] is an internal interface used by The Times to test GenAI. *See* Ex. 2. This tool interfaces with OpenAI's large language models via API, [REDACTED]

Ex. 3. [REDACTED]

See Exs. 4, 2.

On January 30, 2025, OpenAI deposed The Times's Rule 30(b)(6) custodial witness Eric Mandel, who testified that the Times uses a platform called [REDACTED] to log [REDACTED] entries. Ex. 5 (Mandel Tr.) at 128:13–20, 130:5–9, 140:25–141:14; Ex. 18 at 1; *see also* Ex. 14. Mandel's notes indicate that [REDACTED] logs are [REDACTED].

¹ Counsel for The Times subsequently represented that [REDACTED].

██████████.” Ex. 18 at 2. Mandel lacked knowledge regarding any other aspect of ██████████ data retention or deletion. *E.g.*, Ex. 5 at 130:10–15. Indeed, Mandel was wholly unprepared to testify to ██████████. *Id.* at 46:8–18, 146:23–147:7.

OpenAI thereafter sought further discovery into ██████████, including counts for chats from relevant Times employees. Exs. 6, 8. The Times agreed to produce ██████████ data, subsequently revealing the repository contains only ██████████ individual entries ██████████ of total data). Exs. 7, 19. The Times also provided the numbers of entries stored for a subset of Times employees, which belie the accuracy of its prior representations about ██████████ data preservation. Ex. 8.

In particular, The Times revealed that it had ██████████ logs for ██████████ designated custodians. For example, documents reflect that Chris Wiggins and Hannah Poferl—the Times’s Chief Data Scientist and its Chief Data Officer/Assistant Managing Editor, respectively—██████████. *See* Ex. 9. Yet, The Times asserted ██████████. Likewise, there is no question that Zach Seward, The Times’s Editorial Director of AI Initiatives, was directly involved in ██████████. Ex. 4. The Times also retained ██████████ logs for Mr. Seward. The Times also has ██████████ logs stored for Alex Hardiman, Head of Product and ██████████, Rebecca Grossman-Cohen and Sam Felix, ██████████, and Jason Sobel, The Times’s Chief Technology Officer.

OpenAI sought answers regarding these figures, which appear to contradict The Times’s sworn statements about its ██████████ preservation policy. Ex. 10. The Times did not claim that it lacked ██████████ logs for these key individuals because they never or rarely used ██████████. In fact, the Times could neither explain these discrepancies nor provide a date certain for when they would. Nor would they provide counts for additional witnesses. Ex. 11.

2. The Times has conceded the relevance of ██████████ data and has not raised any cognizable burden to preserving it.

Since The Times has not responded to OpenAI’s preservation-related questions, OpenAI cannot know if The Times failed to properly preserve ██████████ data in violation of its own policies and, if so, how much data was deleted. Yet, The Times has acknowledged that this data is relevant and should have been preserved.

The Times has agreed that documents (a) reflecting its use of OpenAI’s LLMs in reporting and creation of content, and (b) related to its A.I. Initiatives program, are relevant. *See* NYT ECF 238 at 2. ██████████ data fits both of these categories. As just two examples, (1) a Times employee ██████████, *see* Ex. 12; and (2) the Times’s A.I. Initiatives team ██████████, *see* Ex. 13. Indeed, The Times’s agreement to produce such documents has served as the basis for orders of this Court. *see* NYT ECF 344 (hereinafter, “November Order”) at 1 n.1; *see also* MDL ECF

82 (hereinafter “May Order”) at 10 (“The Times has already agreed to produce documents relating to its use of defendants’ products.”).²

The Times has also agreed to produce [REDACTED] data specifically. *See* Ex. 15. Moreover, the evidence suggests that the Times began [REDACTED] precisely because it was aware of its preservation duties—it appears The Times [REDACTED] within days of filing this lawsuit in December 2023. *See* Ex. 14. Having conceded³ the relevance of [REDACTED] data, the Times unquestionably had a duty to preserve it. *See Zubulake v. UBS Warburg LLC*, 220 F.R.D. 212, 217 (S.D.N.Y. 2003) (parties “must not destroy unique, relevant evidence that might be useful to an adversary”).

The Times has not argued that preserving [REDACTED] data would be unduly burdensome or otherwise improper. Nor could it. Irrespective of this litigation, The Times’s preservation policy purportedly calls for the preservation of [REDACTED] data for [REDACTED]. *See* Ex. 18. And the [REDACTED] data The Times reportedly has preserved indicates that preservation of relevant data would not be voluminous: the [REDACTED] data that The Times possesses equals [REDACTED] % of the storage capacity of an entry-level iPhone.

3. The Court should order The Times to answer OpenAI’s preservation questions by July 2.

The Times has failed to substantively respond to OpenAI’s preservation questions or provide a date certain by which it will do so. Ex. 11. This is inappropriate, not only because a failure to preserve key evidence is a serious lapse for which further remedies may be necessary, *Zubulake*, 220 F.R.D. at 219–22, but also because The Times has represented that the close of discovery is near. Judge Stein has also directed the parties move this litigation forward. *See* Ex. 17 (May 22, 2025 Hr’g Tr.) at 24:13–16, 36:20–25, 69:16–70:9.

OpenAI asks the Court to compel The Times to, by July 2, (1) confirm it is complying with its own preservation practices for [REDACTED] data, at least as to each relevant Times employee listed in Exhibit 1; (2) provide counts for the number of [REDACTED] logs for each of those employees; and (3) answer OpenAI’s preservation questions, listed in Exhibit 1. This information is necessary so that OpenAI may ascertain if any relevant data was deleted and so the parties can finish meeting and conferring regarding the production of [REDACTED] data.

² To the extent The Times suggests that the portion of the May Order addressing Microsoft’s request for “ChatGPT use by the *Authors Guild* plaintiffs” impacts OpenAI’s request for [REDACTED] data, that is incorrect. May Order at 12. The May Order explicitly held documents reflecting The Times’s use of Defendants’ LLMs for business purposes is relevant. *See id.* at 10. The Court held, by contrast, that Microsoft could not take discovery into *every* ChatGPT prompt by the *Authors Guild* plaintiffs because the *Authors Guild* contributory-infringement claim against Microsoft does not identify end-users as primary infringers. *See id.* at 12.

³ With or without The Times’s concession, [REDACTED] data is relevant to, among other things, OpenAI’s fair use and substantial non-infringing uses defenses.

Respectfully,

KEKER, VAN NEST &
PETERS LLP⁴

/s/ Andrew Bruns
Andrew Bruns

LATHAM & WATKINS
LLP

/s/ Herman H. Yue
Herman H. Yue

MORRISON & FOERSTER
LLP

/s/ Rose S. Lee
Rose S. Lee

⁴ All parties whose electronic signatures are included herein have consented to the filing of this document.